

ОСНОВНЫЕ ЗАДАЧИ СРАВНИТЕЛЬНОГО АНАЛИЗА БЕЛОРУССКОЙ И РУССКОЙ ЛЕКСИКИ ШКОЛЬНЫХ УЧЕБНИКОВ ПО ЯЗЫКУ И ЛИТЕРАТУРЕ

За последние десятилетия проблема создания и использования корпусов текстов различных языков становится общезначимой. Существует большое число научных работ, в которых показано, какую информацию лингвисты, литературоведы, педагоги и методисты могут извлечь из таких корпусов текстов. Так, недавно опубликованы два тома Международного справочника *Corpus Linguistics* [9; 10], содержащие статьи более 60 авторов, показывающие как использовать корпусы текстов в лексикографии, распознавании значений многозначных слов, обучении, диалектологии, реферировании, переводе текстов и т.д.

При этом исследователи ориентировались, как правило, на общепризнанные типы корпусов текстов: исследовательский, иллюстративный, статистический, динамический корпусы и корпус параллельных текстов. Говоря о возможностях использования таких корпусов текстов, известный английский исследователь Г. Уиддоусон отмечает, что анализ даже ограниченного корпуса текстов может выявить модели использования языка, о которых сами носители языка, в том числе и те, кто произвел этот языковой материал, не имеют представления [11, с. 73–74].

Вместе с тем, как отмечают специалисты, уже использовавшие корпусы текстов для проведения различных научных исследований, «возможность проводить собственные наблюдения при обеспечении их статистической достоверности не может не сказаться на организации учебного процесса и на роли в нем студента и преподавателя. Опора на корпус ставит также вопрос о балансе между творчеством и использованием готовых фраз в языке и заставляет задуматься о том, как можно использовать такие методы описания для интеграции семантической, прагматической и культурной информации в учебных материалах». Для этого сейчас создаются учебные корпусы текстов (*Learner Corpora*).

Учебные корпусы текстов можно разделить на 2 группы:

1. Одноязычные учебные корпусы текстов.
2. Двухязычные параллельные учебные корпусы текстов.

К числу первых чаще всего относят корпусы текстов школьников или студентов, изучающих какой-либо язык как иностранный. Такие корпусы текстов созданы в Бельгии (*Corpus of Learner English*), в Финляндии (*English as Lingua Franca in Academic Setting*), в России («Корпус английских текстов Петербургских школьников»), во Франции (*French learner writing corpus*) и в других странах. Подобные корпусы текстов могут быть с успехом использованы для выявления и анализа наиболее распространенных ошибок в изучаемых языках. Необычайно большой объем лингвистического материала, содержащегося в таких корпусах текстов, и воз-

возможности современных информационных технологий обработки текстов значительно облегчат и ускорят процесс создания новой учебной литературы и процесс отбора демонстрационного материала непосредственно для обучения в аудитории или классе.

Очень интересен в этом плане «Корпус английских текстов Петербургских школьников», созданный в Российском государственном педагогическом университете им. А.И. Герцена [4]. Авторами этих текстов являлись школьники 9–11 классов школ Санкт–Петербурга со средним и выше среднего уровня владения английским языком.

При этом жанры отбираемых для корпуса школьных текстов выбирались в соответствии со школьной программой и форматом ЕГЭ по английскому языку. В основном корпус включал два подкорпуса письменных текстов (эссе и личное письмо) и два подкорпуса устных текстов (монолог и диалог).

В процессе обучения белорусскому и русскому языкам, при подготовке учебников и учебных пособий возникает целый ряд лингвистических и методических проблем, связанных со спецификой этих языков. Их решению в определенной степени будет способствовать создаваемый на кафедре информатики и прикладной лингвистики МГЛУ по заданию «Центра исследований белорусской культуры, языка и литературы НАН Беларуси» корпус текстов школьных учебников по белорусскому и русскому языкам.

Этот корпус относится к числу так называемых сравнимых корпусов текстов. Такие корпуса состоят из текстов на разных языках, похожих или сопоставимых в отношении типа текстов, их тематики, способа подачи информации и т.п. Создаваемый корпус текстов включает тексты школьных учебников по белорусскому языку (10 и 11 классы) и белорусской литературе (10 и 11 классы), а также тексты по русскому языку (10 и 11 классы) и русской литературе (10 и 11 классы).

На первом шаге анализа этих текстов было изучено их тематическое содержание. В результате, в текстах школьных учебников по белорусскому и русскому языкам были выделены следующие тематические разделы:

1. Теория языка.
2. Текст о писателе.
3. Текст о поэте.
4. Фрагмент текста (художественный).
5. Фрагмент текста (публицистический).
6. Фрагмент текста (научный).
7. Вопросы.
8. Задания.
9. Контроль задания.
10. Памятка по выполнению задания.
11. Примеры выполнения задания.
12. Методические указания.

Соответственно, в текстах школьных учебников по белорусской и русской литературе выделены такие тематические разделы:

1. Теория литературы.
2. Текст о писателе.
3. Текст о поэте.
4. Текст о произведении.
5. Зарисовка (анализ конкретного произведения).
6. Вопросы.
7. Задания.
8. Вопросы и задания.
9. Вопросы для повторения.
10. Памятка по выполнению задания.

Далее было определено в каждом учебнике число текстов определенной тематики по языку и по литературе.

На следующем этапе проводимого исследования по всем текстам каждой тематики по специальной компьютерной программе были получены частотно-алфавитные словари. В таком словаре для каждой словоформы указывалась абсолютная частота ее употребления во множестве текстов конкретной тематики, а также общее количество абзацев во всем обработанном множестве текстов конкретной тематики, в которых встретилась словоформа.

Затем словоформы всех созданных частотных словарей были тегированы с использованием стандарта тегирования CES (CorpusEncoding-Standart). При этом каждая словоформа получала набор морфологической и структурной информации.

С опорой на тегированные словоформы частотно-алфавитных словарей с помощью специальных компьютерных программ были тегированы словоупотребления всех отобранных для исследования текстов школьных учебников.

С опорой на полученную информацию можно проводить с использованием компьютеров большой комплекс различных исследований, связанных с сравнительно-сопоставительным анализом белорусской и русской лексики. Такая информация важна как с чисто теоретической точки зрения, так и с точки зрения методики обучения [1; 2; 3; 4; 5; 6; 7; 8]. Предполагается, что к создаваемому корпусу текстов будут иметь доступ школьники с помощью специально созданных компьютерных программ.

ЛИТЕРАТУРА

1. Бордович, А.М. Сопоставительный курс русского и белорусского языков. — Минск: 1999.
2. Барысенка, В.У. Руская мова як пасрэднік пры навучанні беларускай мове як замежнай // Контрастивные исследования языков и культур. Материалы Международного круглого стола, 29–30 октября 2012 года. — Минск, 2013. — С. 149–151.
3. Зубов, А.В. Лингвометодические возможности русско-белорусского параллельного корпуса текстов // Русский язык. Исторические судьбы и современность. IV Международный конгресс исследователей русского языка. Труды и материалы. — М., 2010. — С. 516–517.
4. Камшилова, О.Н. Исследовательский потенциал корпуса английских текстов Петербургских школьников: анализ интерязыка // Известия Российского государствен-

ного педагогического университета им. А.И. Герцена. Научный журнал. 2009. № 4. — С. 114–123.

5. Луц, Л.М. Беларуская мова. Прафесійная лексіка. Практукум. — Мінск, 2014.

6. Сопоставительный анализ восточнославянских языков. — Минск, 2014.

7. Нікалаева, В.М. Да пытання аб кантрастыўным даследванні лексічных сістэм рускай і беларускай моў // Беларуская лінгвістыка. — Мінск, 2013. — Вып. 71. — С. 9–19.

8. Яницкая, А.Ю. Выражение эмоционального состояния и чувства радости по-русски по-белорусски // Славянские языки: системно-описательный и социокультурный аспект исследования. Часть 1. — Брест, 2014. — С. 281–284.

9. Corpus Linguistics / An International Handbook. — Berlin. New-York, 2008. — Vol. 1.

10. Corpus Linguistics / An International Handbook. — Berlin. New-York, 2009. — Vol. 2.